

Pemanfaatan AI dalam Analisis Basis Data

Bintang Augri Faris¹, Daniel Ilyasa², Dwi Apriansyah³, M. Rizki Rianto⁴, Uus Firdaus⁵

¹Universitas Djuanda, m.bintangaugryf@gmail.com

²Universitas Djuanda, muhammaddanielilyasa@gmail.com

³Universitas Djuanda, dwiapriansyahh1998@gmail.com

⁴Universitas Djuanda, mrizkirianto116@gmail.com

⁵Universitas Djuanda, uus.firdaus@unida.ac.id

ABSTRAK

Pemanfaatan kecerdasan buatan (artificial intelligence) telah mengalami perkembangan pesat dalam beberapa dekade terakhir. Penggunaan kecerdasan buatan ini telah menyebar ke berbagai aspek, diantaranya dalam Basis Data. Artikel ini membahas bagaimana integrasi kecerdasan buatan dimanfaatkan dalam analisis basis data untuk mendapatkan wawasan yang berharga dari kumpulan data yang kompleks. Penelitian dilakukan dengan metode analisis literatur untuk mendapatkan temuan – temuan penelitian dalam literatur ini.

Kata Kunci: *Artificial Intelligence, Database, AI, Basis Data, Kecerdasan Buatan*

PENDAHULUAN

Dalam era digital yang terus berkembang, data telah menjadi aset yang sangat berharga bagi berbagai industri dan organisasi. Dalam menghadapi kompleksitas data yang terus meningkat, pemanfaatan kecerdasan buatan (AI) dalam analisis basis data menjadi semakin vital. Penelitian ini bertujuan untuk menggali dan menjelaskan berbagai pemanfaatan integrasi kecerdasan buatan dalam konteks analisis basis data. Fokus utama penelitian ini adalah pemahaman mendalam tentang cara teknologi AI dapat diintegrasikan dalam pengelolaan basis data, serta dampaknya terhadap pengambilan keputusan yang lebih cerdas dan efisien.

AI sendiri merupakan sistem atau teknologi yang dirancang untuk melakukan tindakan yang benar atau tepat berdasarkan pengetahuan yang dimilikinya. AI melibatkan proses berpikir dan penalaran, serta perilaku yang sesuai dengan standar manusia atau standar ideal yang disebut sebagai rasionalitas. (Russell, J. S., & Norvig, P. (2020). *Artificial Intelligence A Modern Approach Third Edition*)

Sedangkan, *database* atau basis data merupakan kumpulan file yang disusun secara terstruktur dan saling berelasi, di mana relasi antar file ditunjukkan melalui kunci-kunci yang ada pada tiap file. Dalam konteks perusahaan atau instansi, sebuah *database* mewakili satu kesatuan data yang digunakan. Secara lebih umum, *database* juga dapat diartikan sebagai kumpulan catatan atau data yang tersimpan dalam komputer, sehingga dapat diakses oleh program komputer untuk menjawab pertanyaan atau permintaan informasi. Data yang diambil melalui pertanyaan-pertanyaan ini kemudian menjadi informasi yang berguna untuk pengambilan keputusan. (Firdaus, U. ., Jihan Fachrani, Risdianto Irawan, & Azharudin. (2023). Sistem Informasi Pengelolaan Data Inventaris Berbasis Web. Karimah Tauhid, 2(4), 1257–1265.)

Latar belakang penelitian pemanfaatan ini dipicu oleh beberapa faktor. Pertama, meningkatnya kompleksitas data yang dihasilkan oleh berbagai sumber, seperti sensor, perangkat *IoT (Internet of Things)*, media sosial, transaksi bisnis dan lain sebagainya. Data-data ini seringkali bersifat besar, heterogen, dan cepat berubah, menciptakan tantangan besar dalam mengelola, menganalisis, dan menggali informasi yang bernilai dari data tersebut. Kedua, keberadaan potensi besar kecerdasan buatan, yang mampu mengolah data secara cepat dan akurat, membantu mengidentifikasi pola-pola tersembunyi, serta memberikan wawasan mendalam yang sulit dijangkau oleh metode analisis tradisional.

METODE PENELITIAN

Metodologi penelitian ini adalah dengan melakukan pendekatan analisis dokumen untuk menjelajahi berbagai sumber literatur, termasuk jurna ilmiah, buku dan artikel terkait pemanfaatan kecerdasan buatan dalam analisis data. Pendekatan ini memungkinkan peneliti untuk mengumpulkan data sekunder dari berbagai sumber yang relevan dan otoritatif, serta menggali wawasan yang telah dipublikasikan oleh para pakar dan peneliti dalam bidang kecerdasan buatan dan analisis basis data.

PENGUMPULAN DAN ANALISIS DATA

Penelitian ini melibatkan pemilihan sumber literatur yang relevan dan terkini mengenai integrasi kecerdasan buatan dalam analisis basis data. Sumber-sumber literatur ini mencakup jurnal ilmiah, artikel, dan buku yang dipublikasikan oleh lembaga-lembaga riset. Pemilihan sumber dilakukan berdasarkan kriteria keakuratan dan relevansi informasi yang disajikan.

Data untuk penelitian ini diperoleh melalui pencarian sistematis dalam basis data elektronik yang meliputi database akademik, perpustakaan digital, dan repositori riset. Proses pengumpulan data dilakukan dengan menggunakan kata kunci yang relevan seperti "*Artificial intelligence in database analysis*" "*AI integration in database*" dan sejenisnya. Pencarian data dilakukan melalui mesin pencari ilmiah dan mesin pencari general untuk memastikan akurasi dan keberlanjutan informasi yang ditemukan

HASIL DAN PEMBAHASAN

Setelah melakukan berbagai penelitian literatur, penulis menemukan beberapa pemanfaatan AI dalam integrasinya pada basis data.

1. *Intelligent Database Interface (IDI)*

Intelligent Database Interface (IDI) adalah antarmuka berbasis *cache* yang memberikan sistem Kecerdasan Buatan akses yang efisien ke satu atau lebih basis data pada satu atau lebih *Database Management System (DBMS)* jarak jauh.

Konsep *IDI* mirip dengan antarmuka sistem reguler pada umumnya, *IDI* merupakan antarmuka pengguna yang dapat diakses dan efisien yang bertujuan memberikan akses yang optimal ke berbagai basis data di beberapa sistem pengelolaan basis data (*DBMS*) jarak jauh yang mendukung *Structured Query Language (SQL)*. *Intelligent Database Interface Language (IDIL)* merupakan bahasa kueri unik dari *IDI*, yang berinteraksi dengan *SQL* dengan menerjemahkan kueri menjadi *SQL* dan mengirimkannya ke *DBMS* yang sesuai. Hasil dari antarmuka basis data cerdas ini dikembalikan (*returned*) satu tuple pada satu waktu.

Ada empat komponen utama pada *IDI* yang terdiri dari *Schema Manager*, *DBMS Connection Manager*, *Query Manager* dan *The Cache Manager*.

1) *The Schema Manager (SM)*

Informasi dari basis data yang dideklarasikan ditentukan oleh *SM*. *SM* juga menyediakan informasi untuk relasi-relasi individual dalam basis data kepada manajer kueri. Sebagian dari tanggung jawab manajer skema adalah memproses deklarasi basis data, mengakses, dan menyimpan informasi skema untuk basis data yang dinyatakan, serta mengelola alias nama relasi, terutama saat terdapat nama-nama yang identik di dua atau lebih basis data yang berisi relasi-relasi tersebut. Setiap kali koneksi ke basis data dibuat, manajer skema secara otomatis mengakses daftar nama relasi yang terdapat dalam basis data tersebut. Daftar ini kemudian dicache untuk akses di masa mendatang jika koneksi ditutup dan dibuka kembali suatu saat nanti. Dalam situasi tersebut, manajer skema hanya akan mengakses informasi skema *DBMS* jika secara eksplisit diarahkan untuk melakukannya. Jika tidak, daftar nama relasi yang dicache akan digunakan.

2) *DBMS Connection Manager (DCM)*

DCM bekerja untuk mengatur seluruh koneksi antara basis data dan *DBMS Remote* (*DBMS* jarak jauh). Tugasnya melibatkan proses permintaan untuk membuka dan menutup koneksi ke basis data, serta melaksanakan semua operasi *I/O* pada tingkat dasar yang terkait dengan koneksi tersebut. Dalam lingkungan *IDI*, setiap basis data hanya diizinkan memiliki satu koneksi aktif yang terkait dengannya secara bersamaan. Lebih lanjut, setiap koneksi tidak diperbolehkan memiliki lebih dari satu aliran hasil kueri atau *generator* yang terkait dengannya; hanya satu generator yang boleh aktif pada satu waktu.

3) *Query Manager*

QM bertugas menerjemahkan kueri *IDIL* ke *SQL* dan mengirimkannya ke *DBMS* melalui *DCM*. Jika kueri sukses dieksekusi, *QM* menghasilkan generator untuk hasilnya. *Generator* ini digunakan untuk mendapatkan tuple hasil atau menghentikan kueri. *DCM* mengelola generator karena hasil kueri bisa direpresentasikan dalam bentuk aliran data langsung dari *DBMS* atau elemen *cache*. *QM* hanya mengirimkan permintaan kueri ke *DCM* untuk pengelolaan hasilnya.

4) *Cache Manager (CM)*

CM bertugas mengelola penyimpanan sementara hasil kueri. Tugasnya meliputi pengenalan kueri *IDIL* yang sudah ada hasilnya di dalam *cache*, menyimpan hasil kueri ke dalam *cache*, dan menggantikan elemen-elemen *cache* yang ada.

(McKay, D.P., & Finin, W.T. (n.d) *The Intelligent Database Interface: Integrating AI and Database Systems*) (Unuriode, O. A., & Olalekan, D. M. (2023, Oktober). *THE INTEGRATION OF ARTIFICIAL INTELLIGENCE INTO DATABASE SYSTEM(AI-DB INTEGRATION REVIEW)*)

2. *Text-to-SQL* dengan LLM

LLM merupakan system *Deep Learning* ataupun *Natural Language Processing* yang dirancang khusus untuk memproses bahasa, seperti *GPT-3.5* dan 4 Dengan menggunakan *Neural Network multilayer recurrent, Large Language Models (LLMs)* dirancang untuk menjadi alat kecerdasan buatan (*AI*) yang mampu menghasilkan teks mirip dengan bahasa manusia setelah dilatih dengan volume data yang sangat besar. Maka dari itu LLM ini terbiasa dengan kumpulan – kumpulan teks yang besar, dan tujuan utama dari pelatihan LLM adalah untuk dapat dengan sukses memprediksi kata berikutnya atau kata yang tidak ada dalam suatu kalimat atau paragraf. LLM ini juga dirancang untuk meniru gaya bahasa, mereka mampu menggambarkan dan memproses hampir setiap kalimat dengan cara yang memungkinkannya menghasilkan kelanjutan terprediksi yang sesuai dengan tata bahasa dan bermakna. (Naveed, H., & Khan, U.A.(2023) *A Comprehensive Overview of. Computation and Language*)

Text-to-SQL sendiri merupakan pendekatan yang mengotomatisasi proses penerjemahan bahasa sehari-hari manusia kedalam kueri *SQL* yang dapat dipahami oleh suatu basis data. Dengan *text-to-SQL* dapat mengurangi kebutuhan kemahiran dalam *SQL* sehingga meningkatkan aksesibilitas basis data dan menjembatani antara pengguna yang belum mahir dengan sistem basis data, dengan begitu dapat pula meningkatkan efisiensi dalam pemrosesan data. (Gao, D., Wang, H., & Li, Y. (2023). *Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation*) (Sun, R., O, S., Nakhost, H., & Dai, H. (n.d.). *SQL-PALM: IMPROVED LARGE LANGUAGE MODEL.*

Pada eksperimen yang dilakukan dalam (Li, J., Hui, B., & Qu, G. (2023). *Can LLM Already Serve as A Database Interface? A Big Bench for Large-Scale Database Grounded Text-to-*

SQL) memberikan hasil perbandingan antara beberapa Model dari *LLM* dengan Performa Manusia, bahwa Pengembangan Data dengan GPT-4 + DIN-SQL dengan pengetahuan memberikan akurasi sebesar 58.79% sedangkan Performa Manusia bisa memberikan tingkat akurasi hingga 90.27%. Untuk saat ini model *LLM* terbaik seperti GPT-4 pun masih belum memberikan hasil yang cukup, akan tetapi dengan adanya eksperimen - eksperimen semacam ini akan memberikan tingkat kemajuan *LLM* dalam integrasi basis data.

Disamping itu, penerapan *LLM* dalam penelitian basis data dapat menangani masalah basis data seperti penemuan data (*data discovery*), pembersihan data (*data cleansing*) dan integrasi data. Kueri data dengan *Chat GPT* sendiri dapat mengurangi ketergantungan akan kueri-kueri *SQL* dan meningkatkan efisiensi pada pengambilan data dan analisis beberapa aplikasi tertentu. (Yahia, A. S., & Bonifati, A. (2023). From Large Language Models to Databases and Back. A discussion on research and education)

KESIMPULAN

Dalam rangka menghadapi kompleksitas data yang terus meningkat, penelitian ini menjelaskan pemanfaatan integrasi kecerdasan buatan (AI) dalam analisis basis data. Fokus utama penelitian adalah pemahaman mendalam tentang bagaimana teknologi AI dapat diintegrasikan dalam pengelolaan basis data dan dampaknya terhadap pengambilan keputusan yang lebih cerdas dan efisien.

Hasil penelitian menunjukkan bahwa *Intelligent Database Interface (IDI)* merupakan salah satu pemanfaatan AI yang signifikan dalam pengelolaan basis data. *IDI*, sebagai antarmuka berbasis *cache*, memberikan akses efisien ke basis data melalui *Database Management System (DBMS)* jarak jauh. *IDI* menggunakan *Intelligent Database Interface Language (IDIL)* untuk berinteraksi dengan *SQL* dan mengembalikan hasil satu *tuple* pada satu waktu. Komponen utama *IDI*, seperti *Schema Manager (SM)*, *DBMS Connection Manager (DCM)*, *Query Manager (QM)*, dan *Cache Manager (CM)*, berperan dalam mengoptimalkan proses tersebut.

Selain itu, penelitian ini juga mencatat peran penting *Large Language Models (LLMs)*, seperti *GPT-3.5*, dalam memproses kueri dalam *text-to-SQL* yang dapat menerjemahkan bahasa sehari-hari manusia kedalam bahasa yang digunakan pada kueri

Kesimpulan ini diakhiri dengan merinci batasan dan kontribusi penelitian. Batasan penelitian melibatkan ruang lingkup tertentu dalam penggunaan *AI* dalam analisis basis data, sedangkan kontribusinya mencakup pengembangan konsep *IDI* dan pemanfaatan *LLMs* pada *Text-to-SQL* dalam konteks analisis data.

Dengan demikian, penelitian ini memberikan wawasan tentang peran vital kecerdasan buatan dalam meningkatkan pengelolaan basis data dan pengambilan keputusan. Sebagai saran untuk penelitian selanjutnya, disarankan untuk mengeksplorasi lebih lanjut aplikasi *AI* dalam skenario bisnis dan industri tertentu. Penelitian lebih lanjut dapat juga melibatkan pengembangan teknologi *AI* yang lebih canggih untuk meningkatkan efisiensi analisis data.

REFERENSI

- Ahmad , K., & Iqbal, W. (2023). Data-Driven Artificial Intelligence in Education:.
Division of Information and Computing Technology, College of Science and Engineering, 17, 12 - 31. doi:10.1109/TLT.2023.3314610
- Li, J., Hui, B., & Qu, G. (2023). Can LLM Already Serve as A Database Interface? A
BIG Bench for Large-Scale Database Grounded Text-to-SQLs.
- Naveed, H., & Khan, U. A. (2023). A Comprehensive Overview of. *Computation and Language. doi:https://doi.org/10.48550/arXiv.2307.06435*
- Russell , J. S., & Norvig, P. (2020). *Artificial Intelligence A Modern Approach Third Edition.*

- Singh, C. (2023). Machine Learning in Pattern Recognition. *European Journal of Engineering and Technology Research*, 8(2), 63-68.
doi:10.24018/ejeng.2023.8.2.3025
- Sun, R., Ö, S., Nakhost, H., & Dai, H. (n.d.). SQL-PALM: IMPROVED LARGE LANGUAGE MODEL. doi:<https://doi.org/10.48550/arXiv.2306.00739>
- Gao, D., Wang, H., & Li, Y. (2023). Text-to-SQL Empowered by Large Language Models: A.
- McKay, D. P., & Finin, W. T. (n.d.). The Intelligent Database Interface: Integrating AI and Database Systems.
- Rahmani, A. M., & Azhir, E. (2021). Artificial intelligence approaches and mechanisms for big data analytics: a systematic study.
doi:<https://doi.org/10.7717/peerj-cs.488>
- Reder, G. K., & Gower, A. H. (2023). Genesis-DB: a database for autonomous laboratory. *Databases and ontologies*, 3(1). doi:10.1093/bioadv/vbad102
- Unuriode, O. A., & Olalekan, D. M. (2023, Oktober). THE INTEGRATION OF ARTIFICIAL INTELLIGENCE INTO DATABASE SYSTEM(AI-DB INTEGRATION REVIEW). *International Journal on Cybernetics & Informatics*, Vol 12(6). doi:10.5121/ijci.2023.1206012
- Wang, X., & Zhang, L. (2021). Artificial intelligence and database technology work together. *Journal of Physics: Conference Series*. doi:10.1088/1742-6596/1955/1/012004
- Yahia, A. S., & Bonifati, A. (2023). From Large Language Models to Databases and Back. doi: <https://doi.org/10.48550>
- Firdaus, U. ., Jihan Facrani, Risdianto Irawan & Bonifati, Azharudin. (2023). Sistem Informasi Pengelolaan Data Inventaris Berbasis Web. doi: <https://doi.org/10.30997/karimahtauhid.v2i4.10032>